TRANSFAC® release 2017.3

The TRANSFAC® database on transcription factors, their genomic binding sites and DNA-binding motifs (PWMs), contains these new data features:

- **Annotation of transcription factor binding sites based on sequence conservation**

  Known transcription factor binding sites located in human, mouse or rat genomes were extracted from TRANSFAC® and highly conserved sites were retained. Given high conservation as a prerequisite, binding sites were annotated for the two other species in respective genomic location if not more than one mismatch was observed in the sequence alignment with the primary species. This resulted in 1,565 new binding site entries.

- **ChIP-Seq experiment browse pages**

  New browse page for 161 human DNase hypersensitivity ChIP-Seq experiments imported from ENCODE. The genomic intervals for each data set can be downloaded in .BED format. The TFBS and DNase ChIP-Seq experiment browse pages can be accessed from the tools menu. The BED download buttons have also been added to the TFBS ChIP-Seq experiment browse page, providing easier access for the now 2,032 data sets.

  113 new transcription factor binding site ChIP-Seq experiments released by the ENCODE phase 3 project between February 2017 and May 2017 https://www.encodeproject.org/matrix/?type=Experiment&status=released. The data sets comprise 1,329,758 fragments bound by 98 distinct transcription factors, of which 66 factors were not yet covered by ChIP-Seq data. For 71 of the sets, an existing positional weight matrix for the respective transcription factor was used together with the MATCH tool to predict altogether 816,574 best binding sites inside the fragments. Predicted best binding sites as well as complete fragments are available in FASTA and BED format via the ChIP Experiment Reports, as are lists of genes in a distance range to the fragments as specified by the user.

- **Reorganization of the in vivo transcription factor bound fragment section on a Locus Report**

  To improve clarity, only those fragment are listed that overlap with the best supported promoter of the entry and as additional information the table contains now the relative position of the fragment to the transcription start site (TSS) of the promoter, as well as the sequence of the predicted best binding site for the transcription factor inside the fragment.

- **HOCOMOCO v10 matrix library integration**

  134 mononucleotide position weight matrices based on ChIP-Seq experiments have been incorporated from HOCOMOCO v10 (http://hocomoco.autosome.ru/).

- **Human SNP content update**

  The new March 2017 dbSNP Build 150 data for human has been integrated and increases the number of SNPs mapped to human promoter sequences more than two-fold from 34,839,288 in the last release to 73,423,232.

- **Ensembl version update**

  Genomic information for genes, promoters, and ChIP fragments for the species human, mouse, rat, macaque, and Arabidopsis is now based on Ensembl release 89.

PROTEOME™ (HumanPSD™+TRANSPATH®) release 2017.3

The Human Proteome Survey Database (HumanPSD™) with focus on human proteins as disease biomarkers and drug targets contains these new data features:

- **Integration of new clinical trial (CT) data sources**

  Integration of new data on clinical trials from clinicaltrials.gov and OpenTrials (https://opentrials.net/), covering studies from, among others, European, Japanese and Australian registries. The number of CT-Disease-Drug assignments increases from 227,170 to 349,417, while the CT-Disease assignments increase from 316,785 to 641,872.

- **Improved user data management**

  The "storage" link in the "my data" menu loads an overview page with usage space statistics for each stored user data file or result list, which can also be sorted by storage space or alphabetically, for easier management or deletion of obsolete files. (This feature is also available in TRANSFAC®).

- **Quick search for disease and drug entries**

  The quick search menu includes now options to search for diseases or drugs by external identifiers, such as MeSH ID, Drugbank ID, or Pubchem CID.

- **Link-out to BRENDA professional - the comprehensive enzyme information system**

  Locus reports of genes/proteins with enzymatic function now contain links to BRENDA, which can be accessed by users with a valid BRENDA subscription.

The TRANSPATH® database on mammalian signal transduction and metabolic pathways contains these new data features:

- **New phosphorylation targets content**

  1,000 new phosphorylation reactions have been added, describing substrates and their phosphosites for key kinases such as ERK1, SYK, Plk1, and Aurora-A/B.